PATENT
Attorney Docket No. YOR920020018US1

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of	
Sankar BASU et al.	Examiner: SMITH, Peter J.
Serial No: 10/056,546	Art Unit: 2176
Filed: January 24, 2002	
For: A METHOD AND APPARATUS FOR ACTIVE ANNOTATION OF MULTIMEDIA CONTENT	

APPEAL BRIEF

Board of Patent Appeals and Interferences
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

The applicant submits this brief pursuant to 37 C.F.R. §41.37(a)(1) in furtherance of the Notice of Appeal filed December 21, 2006, setting a two-month shortened statutory period of brief filing expiring February 21, 2007.

Please charge Deposit Account 50-0510 the $500 fee for filing this Appeal Brief. No other fee is believed due with this Appeal Brief, however, should another fee be required please charge Deposit Account 50-0510.

Real Party In Interest

The real party in interest is International Business Machines Corporation.

Related Appeals And Interferences

None.

Status of Claims

Claims 1-28 are pending in the present Application, with claims 1, 16 and 22 being independent claims.

The rejection of claims 1-28 is appealed. The table below summarizes the status of the claims.

Claim(s)	Status	Appealed
1	Amended	Yes
2-4	Original	Yes
5	Amended	Yes
6-15	Original	Yes
16	Amended	Yes
17-19	Original	Yes
20	Amended	Yes
21	Original	Yes
22-28	Amended	Yes

Status of Amendments

No amendments to the claims were made after the Final Office Action dated September 21, 2006 ("FOA").

Summary of the Claimed Subject Matter

The present invention relates to efficient interactive annotation or labeling of multimedia content to facilitate effective searching, filtering and usage of content. App., pg. 1, ln. 4-6. In one embodiment of the invention, efficiency is realized by prompting a user to annotate only a small set of selected example content, with the results propagated to the annotation of a maximum number of other multimedia content. App.,

pg. 4, ln. 11-14, pg. 7, ln. 2-4. This causes the user to annotate as few examples as possible. App., pg. 5, ln. 7-8.

Claim 1 recites a method for generating persistent annotations of multimedia content. App. pg. 6, ln. 16-18 and Fig. 1, items 101-103. A selecting step actively selects examples of multimedia content to be annotated by a user. App. pg. 6, ln. 25-29 and Fig. 2, item 101. The examples of multimedia content are selected based on at least one criterion for achieving a maximal disambiguation result such that only those examples which are most ambiguous are selected. App. pg. 7, ln. 2-6 and Fig. 3, items 301 and 202.

An accepting operation accepts input annotations from the user for the selected examples. App. pg. 7, ln. 21-27 and Fig. 4, item 501. A propagating operation propagates the input annotations to other instances of multimedia content. App. pg. 8, ln. 3-10 and Fig. 5, item 802. A storing operation stores the input annotations and the propagated annotations. App. pg. 6, ln. 21-22 and Fig. 1, item 103.

Claim 16 recites a system for generating persistent annotations of multimedia content. App. pg. 6, ln. 16-18 and Fig. 1, items 101-103. The system includes means for actively selecting examples of multimedia content to be annotated by a user. App. pg. 6, ln. 25-29 and Fig. 2, item 101. The examples of multimedia content are selected based on at least one criterion for achieving a maximal disambiguation result such that only those examples which are most ambiguous are selected. App. pg. 7, ln. 2-6 and Fig. 3, items 301 and 202.

The system further includes means for accepting input annotations from the user for the selected examples (App. pg. 7, ln. 21-27 and Fig. 4, item 501), means for propagating the input annotations to other instances of multimedia content (App. pg. 8, ln. 3-10 and Fig. 5, item 802) and means for storing the input annotations and the propagated annotations (App. pg. 6, ln. 21-22 and Fig. 1, item 103).

Claim 22 recites a computer program product in a computer readable medium for generating persistent annotations of multimedia content. App. pg. 6, ln. 16-18, pg. 15, ln. 12-22 and Fig. 1, items 101-103. The computer program product includes instructions for performing one or more repetitions of actively selecting of examples of multimedia content to be annotated by a user. App. pg. 6, ln. 25-29 and Fig. 2, item 101. The examples of multimedia content are selected based on at least one criterion for achieving a maximal disambiguation result such that only those examples which are most ambiguous are selected. App. pg. 7, ln. 2-6 and Fig. 3, items 301 and 202.

The computer program product further includes instructions for performing one or more repetitions of accepting input annotations from the user for the selected examples (App. pg. 7, ln. 21-27 and Fig. 4, item 501), propagating the input annotations to other instances of multimedia content (App. pg. 8, ln. 3-10 and Fig. 5, item 802) and storing the input annotations and the propagated annotations (App. pg. 6, ln. 21-22 and Fig. 1, item 103).

Grounds for Rejection to be Reviewed on Appeal

I. Claims 1-23, 25 and 27 were rejected under 35 USC §103 as unpatentable over U.S. Patent No. 6,804,684 issued to Stubler et al. (hereinafter "Stubler") in view of U.S. Patent No. 5,253,362 to Nolan (hereinafter "Nolan").

II. Claims 24, 26 and 28 were rejected under 35 USC §103 as unpatentable over Stubler in view of Nolan and further in view of U.S. Patent No. 6,816,847 to Toyama (hereinafter "Toyama").

Argument

I. CLAIMS 1-23, 25 AND 27 ARE PATENTABLE UNDER 35 USC §103 OVER STUBLER IN VIEW OF NOLAN

Claim 1

In rejecting claims under U.S.C. §103, the Examiner bears the initial burden of establishing a *prima facie* case. In re Oetiker, 977 F.2d 1443, 1445, 24 USPQ 1443, 1444 (Fed. Cir. 1992). To establish *prima facie* obviousness of a claimed invention, all the claim limitations must be taught or suggested by the prior art. In re Royka, 490 F.2d 981, 180 USPQ 580 (CCPA 1974). "All words in a claim must be considered in judging the patentability of that claim against the prior art." In re Wilson, 424 F.2d 1382, 1385, 165 USPQ 494, 496 (CCPA 1970).

Independent claim 1 recites, in part, "actively selecting examples of multimedia content to be annotated by a user, wherein the examples of multimedia content are selected based on at least one criterion for achieving a maximal disambiguation result, such that only those examples which are most ambiguous are selected."

In rejecting claim 1, the Office Action argues that although Stubler does not teach actively selecting examples of multimedia content to be annotated by a user, wherein the examples of multimedia content are selected based on at least one criterion for achieving a maximal disambiguation result, this teaching is found in Nolan. FOA, pg. 3-4. The Office Action cites column 5, lines 50-56 and column 6, lines 13-21 of Nolan in support of its position. FOA, pg. 4.

Before addressing the specific passages cited by the Examiner, it may be helpful to broadly discuss the teachings of Nolan. Nolan relates to an automated records keeping management system, such as in a hospital based patient record keeping system. Nolan, col. 1, ln. 57-60.

Fig. 1 of Nolan shows a patient information screen envisioned by the inventors. Nolan, col. 4, ln. 62-64. A nurse or other user can enter patient information by selecting data cell and selecting various menu options. Nolan, col. 5, ln. 14-19 and Fig. 3. For example, by selecting data cell 465 in Fig. 3, a user can enter comments associated with the patient's blood pressure reading at 11:30 by selecting "(4) NURSING ANNOTATION" from a pop-up menu. Nolan, col. 5, ln. 20-27 and Fig. 4. Nolan also contemplates a list of predefined, commonly used annotations that a user may select from a selection menu. Nolan, col. 5, ln. 51-56 and Fig. 6.

Turning to specific passages in Nolan cited by the Examiner, column 5, lines 50-56 and column 6, lines 13-21 state:

> Additionally, when an annotation is to be entered, such as in FIG. 6, a window 498 may be displayed. Window 498 provides a list of predefined, commonly used annotations. By selecting one of the annotations from window 498, that annotation will automatically be placed into annotation field 484 of window 480. Nolan, col. 5, ln. 50-56.
>
> In FIG. 8, a second embodiment of the present invention is illustrated. This embodiment provides a window 110 of predefined, commonly used nursing annotations for the nurse to select from. With window 110, the nurse may select an option, such as option 1 and have it automatically transfered to nursing progress notes 100. The date, shift, and time would then need to be entered for the annotation. The RN entry for the annotation may either be made by the nurse or automatically made by the system. Nolan, col. 6, ln. 13-21.

The Examiner takes the position that "Nolan teaches a list predefined commonly used annotation for which the nurse can select from. Using the broadest interpretation of Nolan's teaching, the Examiner concludes that the list of annotations would inherently be comprised of annotation examples, which are the most ambiguous." FOA, pg. 4.

The Appellants respectfully disagree with the Examiner's rational. First, it is clear that Nolan is directed toward a records keeping management system that provides helpful information for interpreting patient records. Rather than

providing annotation examples that are the most ambiguous, the annotations would be inherently as unambiguous as possible since clarity in patient records is very important within the health care industry. Furthermore, the Examiner provides no basis in fact or technical reasoning to reasonably support the determination that the list of annotations would inherently be comprised of annotation examples, which are the most ambiguous, necessarily flows from the teachings of Nolan. In relying upon the theory of inherency, the Examiner must provide a basis in fact and/or technical reasoning to reasonably support the determination that the allegedly inherent characteristic necessarily flows from the teachings of the applied prior art. MPEP 2112 citing Ex parte Levy, 17 USPQ2d 1461, 1464 (Bd. Pat. App. & Inter. 1990) (emphasis in original).

More importantly, claim 1 does not recite that the selected annotations are the most ambiguous, but rather the selected examples multimedia content to be annotated are the most ambiguous. The Appellants respectfully submit that there is no teaching, either in the citation offered by the Examiner or elsewhere in Nolan, of such a limitation.

Obviousness cannot be established by combining prior art to produce the claimed invention absent some teaching or suggestion supporting the combination. In re Fritch, 972 F.2d 1260, 1266, 23 USPQ2d 1780, 1783-84 (Fed. Cir. 1992). The mere fact that the prior art may be modified in the manner suggested by an Examiner does make the modification obvious unless the prior art suggested the desirability of the modification. Id.

The Final Office Action also argues, "It would have been obvious to a person of ordinary skill in the art to combine Stubler with Nolan's teaching of providing a list of commonly used annotation examples for the benefit of achieving a maximal disambiguation result such that only those examples which are most ambiguous are selected." FOA, pg. 4 (emphasis in original).

In the present case, the advantage alleged by the Examiner to justify the proposed combination of Stubler and Nolan does not stand up to close scrutiny. More particularly, the examiner has not explained, and it not evident, why a person of ordinary skill in the art would have found it obvious to reconstruct Stubler to achieve a maximal disambiguation result such that only those examples which are most ambiguous are selected. In this regard, neither Stubler nor Nolan express any appreciation of the efficient interactive annotation or labeling of multimedia content to facilitate effective searching, filtering and usage of content attributed in the Appellant's specification. App., pg. 1, ln. 4-6.

In this light, it is apparent that the only suggestion for combining Stubler and Nolan in the manner advanced by the Examiner stems from hindsight knowledge impermissibly derived from the Applicant's disclosure.

The Appellants respectfully assert that the Examiner has therefore not established a *prima facie* case of unpatentability for claim 1. The Appellants submit that the rejection of claim 1 is improper and should be reversed by the honorable Board.

Claims 2-15

If an independent claim is nonobvious under 35 U.S.C. 103, then any claim depending therefrom is nonobvious. In re Fine, 837 F.2d 1071, 5 USPQ2d 1596 (Fed. Cir. 1988).

Claims 2-15 and 23 are dependent on and further limit claim 14. Since the rejection of claim 1 is believed improper, the rejections of claims 2-15 and 23 are also believed improper for at least the same reasons as claim 1.

Claim 23

Claim 23 is dependent on claim 1 and recites, "The method of claim 1, wherein the at least one criterion includes an ambiguity level of the selected examples." In rejecting claim 23, the Final

Office Action argues that although Stubler does not teach wherein at least one criterion include an ambiguity level of selected examples, such a teaching is found at column 5, lines 50-56 and column 6, lines 13-21 of Nolan. FOA, pg. 7.

Specifically, the Examiner argues, "Nolan teaches a list predefined commonly used annotation for which the nurse can select from. Using the broadest interpretation of Nolan's teaching, the Examiner concludes that the list of annotations would inherently be comprised of annotation examples, which are the most ambiguous." FOA, pg. 7. The Appellants respectfully disagree with the Examiner's interpretation of Nolan.

The cited passages of Nolan (reproduced in the discussion of claim 1 above) make no mention of an ambiguity level of the selected examples. The Examiner provides no basis in fact or technical reasoning to reasonably support the determination that at least one criterion includes an ambiguity level of selected examples, necessarily flows from the teachings of Nolan. In relying upon the theory of inherency, the Examiner must provide a basis in fact and/or technical reasoning to reasonably support the determination that the allegedly inherent characteristic <u>necessarily</u> flows from the teachings of the applied prior art. MPEP 2112 citing <u>Ex parte Levy</u>, 17 USPQ2d 1461, 1464 (Bd. Pat. App. & Inter. 1990) (emphasis in original).

As mentioned above, Nolan contemplates a list of predefined, commonly used annotations that a user may select from a selection menu. Nolan, col. 5, ln. 51-56 and Fig. 6. There is no teaching or suggestion of an ambiguity level of selected examples. Furthermore, the ambiguity level of selected examples found in claim 24 refers back to the selected examples of multimedia content, not to selected annotations as argued by the Examiner. The Appellants therefore respectfully submit that there is no teaching, either in the citation offered by the Examiner or elsewhere in Nolan, of such a limitation.

For at least these reasons, and the reasons given for claim 1, the rejection of claim 23 is believed improper. The Appellants respectfully request reversal of the rejection of claims 23 by the honorable Board.

Claims 16 and 22

Independent claims 16 and 22 recite similar limitations as claim 1 and were rejected under the same grounds as claim 1. FOA, pg. 3-4. Since the rejection claim 1 is believed improper for at least the reasons discussed above, the rejections of claims 16 and 22 are also believed improper for at least the same reasons. The Appellants therefore respectfully submit that the honorable Board should reverse the rejections of claims 16 and 22.

Claims 17-21

Claims 17-21 are dependent on and further limit claim 16. Since the rejection of claim 16 is believed improper, the rejections of claims 17-21 are also believed improper for at least the same reasons as claim 16.

Claims 25 and 27

Claims 25 and 27 recite similar limitations as claim 24 and were rejected under the same grounds as claim 23. FOA, pg. 7. Since the rejection claim 23 is believed improper for at least the reasons discussed above, the rejections of claims 25 and 27 are also believed improper for at least the same reasons. The Appellants therefore respectfully submit that the honorable Board should reverse the rejections of claims 25 and 27.

II. CLAIMS 24, 26 AND 28 ARE PATENTABLE UNDER 35 USC §103 OVER STUBLER IN VIEW OF NOLAN AND FURTHER IN VIEW OF TOYAMA

Claim 24

Claim 24 recites, "The method of claim 1, wherein the at least one criterion includes a confidence level of the selected examples, the confidence level being inversely proportional to a

distance of a new feature of the selected examples from a separating hyperplane in an induced higher dimensional feature space." In rejecting claim 24, the Office Action argues that although neither Stubler nor Nolan teach wherein the at least one criterion includes a confidence level of the selected examples, the confidence level being inversely proportional to a distance of a new feature of the selected examples from a separating hyperplane in an induced higher dimensional feature space, such a teaching is found at column 5, line 15 to column 6, line 50 of Toyama. The Appellants respectfully disagrees with the Examiner's interpretation of Toyama.

The Applicants find only two sentences that deal with selecting examples in the citation provided by the Examiner. The first sentence states, "For example, the images may include a set of web pages, a set of scanned-in pictures, a set of created pictures, a set of drawings, a set of page layouts, etc." The second sentence states, "The set of images in the training set desirably includes a wide variety of images, both those considered aesthetically pleasing, and those considered aesthetically poor." Neither sentence mentions or suggests wherein the at least one criterion includes a confidence level of the selected examples, the confidence level being inversely proportional to a distance of a new feature of the selected examples from a separating hyperplane in an induced higher dimensional feature space.

The Examiner further urges that Toyama discloses the user of support vector machines at column 6, lines 10-50. The specific mention of Support Vector Machines in Toyama states,

> For example, Support Vector Machines build classifiers by identifying a hyperplane that separates a set of positive and negative examples with a maximum margin. In the linear form of SVM that is employed in one embodiment, the margin is defined by the distance of the hyperplane to the nearest positive and negative cases for each class. Maximizing the margin can be express as an optimization problem and search and optimization thus lay at the core of different SVM-based training methods. A post-processing procedure described in the Platt reference is used that employs regularized maximum likelihood fitting to produce estimations of posterior

> probabilities. The method fits a sigmoid to the score that is output by the SVM classifier. Toyama, col. 6, ln. 19-31.

The Applicants respectfully submit that this passage relates to using Support Vector Machines (see http://en.wikipedia.org/wiki/Support_vector_machines attached herewith) for creating functions from a set of labeled training data. The passage contains no teaching or suggestion of using a confidence level inversely proportional to a distance of a new feature of the selected examples from a separating hyperplane in an induced higher dimensional feature space.

For at least these reasons, and the reasons given for claim 1, the rejection of claim 24 is believed improper. The Appellants respectfully request reversal of the rejection of claims 23 by the honorable Board.

Claims 26 and 28

Claims 26 and 28 recite similar limitations as claim 24 and were rejected under the same grounds as claim 24. FOA, pg. 8-9. Since the rejection claim 24 is believed improper for at least the reasons discussed above, the rejections of claims 26 and 28 are also believed improper for at least the same reasons. The Appellants therefore respectfully submit that the honorable Board should reverse the rejections of claims 26 and 28.

Conclusion

In view of the foregoing, Appellant submits that the rejections of Claims 1-9 and 14-21 are improper and respectfully requests that the rejections of Claims 1-9 and 14-21 be reversed by the Board.

Respectfully submitted,

Dated: February 21, 2007

Ido Tuchman, Reg. No. 45,924
Law Office of Ido Tuchman
82-70 Beverly Road
Kew Gardens, NY 11415
Telephone (718) 544-1110
Facsimile (718) 544-8588

Claims Appendix

Claim 1. (previously presented) Method for generating persistent annotations of multimedia content, comprising one or more repetitions of the following steps:

actively selecting examples of multimedia content to be annotated by a user, wherein the examples of multimedia content are selected based on at least one criterion for achieving a maximal disambiguation result such that only those examples which are most ambiguous are selected;

accepting input annotations from said user for said selected examples;

propagating said input annotations to other instances of multimedia content; and

storing said input annotations and said propagated annotations.

Claim 2. (original) The method of claim 1, wherein the step of actively selecting is performed using a selection technique selected from the group consisting of: deterministic and probabilistic.

Claim 3. (original) The method of claim 2, wherein the step of actively selecting, which is performed deterministically or probabilistically, is based on explicit models and feature proximity/similarity measures, and returns one or more examples of multimedia content to be annotated.

Claim 4. (original) The method of claim 2, wherein the step of actively selecting, which is performed deterministically or probabilistically, is based on implicit models and feature proximity/similarity measures, and returns one or more examples of multimedia content to be annotated.

Claim 5. (previously presented) The method of claim 1, wherein an optimization criterion for active selection includes one or more criteria selected from the group consisting of: information measures and confidence.

Claim 6. (original) The method of claim 1, wherein the multimedia content comprises one or more types selected from the group consisting of: images, audio, video, graphics, text, multimedia, Web

pages, time series data, surveillance data, sensor data, relational data, and XML data.

Claim 7. (original) The method of claim 1, wherein the input annotations are created by a user with reference to a vocabulary.

Claim 8. (original) The method of claim 7, wherein the vocabulary contains one or more items selected from the group consisting of: terms, concepts, labels, and annotations.

Claim 9. (original) The method of claim 1, wherein the process of creating input annotations by the user involves multimodal interaction with the user using graphical, textual, and/or speech interface.

Claim 10. (original) The method of claim 1, wherein the input annotations are created by means of steps selected from the group consisting of: creating new annotations, deleting existing annotations, rejecting proposed annotations, and modifying annotations.

Claim 11. (original) The method of claim 7, wherein the vocabulary is adaptively or dynamically organized and/or limited by the system or the user.

Claim 12. (original) The method of claim 9, wherein the multimodal interaction involves speech recognition, gaze detection, finger pointing, expression detection, and/or effective computing methods for sensing a user's state.

Claim 13. (original) The method of claim 1, wherein the determination of the propagation of annotations is made deterministically or probabilistically and on the use of models for each annotation or for joint annotations.

Claim 14. (original) The method of claim 2, wherein the models are created or learned automatically or semi-automatically and/or are updated adaptively from interaction with the user.

Claim 15. (original) The method of claim 2, wherein the models are based on nearest neighbor voting or variants, parametric or

statistical models, expert systems, rule-based systems, or hybrid techniques.

Claim 16. (previously presented) System for generating persistent annotations of multimedia content, comprising:

means for actively selecting examples of multimedia content to be annotated by a user, wherein the examples of multimedia content are selected based on at least one criterion for achieving a maximal disambiguation result such that only those examples which are most ambiguous are selected;

means for accepting input annotations from said user for said selected examples;

means for propagating said input annotations to other instances of multimedia content; and

means for storing said input annotations and said propagated annotations.

Claim 17. (original) The system of claim 16 wherein the means for actively selecting uses a selection technique selected from the group consisting of: deterministic and probabilistic.

Claim 18. (original) The system of claim 17, wherein the means for actively selecting, which uses a deterministic or probabilistic technique, is based on explicit models and feature proximity/similarity measures, and returns one or more examples of multimedia content to be annotated.

Claim 19. (original) The system of claim 17, wherein the means for actively selecting, which uses a deterministic or probabilistic technique, is based on implicit models and feature proximity/similarity measures, and returns one or more examples of multimedia content to be annotated.

Claim 20. (previously presented) The system of claim 16, wherein an optimization criterion for active selection includes one or more criteria selected from the group consisting of: information measures and confidence.

Claim 21. (original) The system of claim 16, wherein the multimedia content comprises one or more types selected from the

group consisting of: images, audio, video, graphics, text, multimedia, Web pages, time series data, surveillance data, sensor data, relational data, and XML data.

Claim 22. (previously presented) A computer program product in a computer readable medium for generating persistent annotations of multimedia content, the computer program product comprising instructions for performing one or more repetitions of the following steps:

actively selecting of examples of multimedia content to be annotated by a user, wherein the examples of multimedia content are selected based on at least one criterion for achieving a maximal disambiguation result such that only those examples which are most ambiguous are selected;

accepting input annotations from said user for said selected examples;

propagating said input annotations to other instances of multimedia content; and

storing said input annotations and said propagated annotations.

Claim 23. (previously presented) The method of claim 1, wherein the at least one criterion includes an ambiguity level of the selected examples.

Claim 24. (previously presented) The method of claim 1, wherein the at least one criterion includes a confidence level of the selected examples, the confidence level being inversely proportional to a distance of a new feature of the selected examples from a separating hyperplane in an induced higher dimensional feature space.

Claim 25. (previously presented) The system of claim 16, wherein the at least one criterion includes an ambiguity level of the selected examples.

Claim 26. (previously presented) The system of claim 16, wherein the at least one criterion includes a confidence level of the selected examples, the confidence level being inversely proportional to a distance of a new feature of the selected examples from a separating hyperplane in an induced higher dimensional feature space.

Claim 27. (previously presented) The computer program product of claim 22, wherein the at least one criterion includes an ambiguity level of the selected examples.

Claim 28. (previously presented) The computer program product of claim 22, wherein the at least one criterion includes a confidence level of the selected examples, the confidence level being inversely proportional to a distance of a new feature of the selected examples from a separating hyperplane in an induced higher dimensional feature space.

Evidence Appendix

Exhibit 1: http://en.wikipedia.org/wiki/Support_vector_machines

Related Proceedings Appendix

None.

Exhibit 1

Support vector machine

From Wikipedia, the free encyclopedia
(Redirected from Support vector machines)

Support vector machines (SVMs) are a set of related supervised learning methods used for classification and regression. Their common factor is the use of a technique known as the "kernel trick" to apply linear classification techniques to non-linear classification problems.

Contents

Linear classification

Motivation

Suppose we want to classify some data points into two classes. Often we are interested in classifying data as part of a machine-learning process. These data points may not necessarily be points in $\mathbf{R}^2$ but may be multidimensional $\mathbf{R}^p$ (statistics notation) or $\mathbf{R}^n$ (computer science notation) points. We are interested in whether we can separate them by a hyperplane. As we examine a hyperplane, this form of classification is known as linear classification. We also want to choose a hyperplane that separates the data points "neatly", with maximum distance to the closest data point from both classes -- this distance is called the *margin*. We desire this property since if we add another data point to the points we already have, we can more accurately classify the new point since the separation between the two classes is greater. Now, if such a hyperplane exists, the hyperplane is clearly of interest and is known as the *maximum-margin hyperplane* or the *optimal hyperplane*, as are the vectors that are closest to this hyperplane, which are called the *support vectors*.

Formalization

We consider data points of the form: $\{(\mathbf{x}_1, c_1), (\mathbf{x}_2, c_2), \ldots, (\mathbf{x}_n, c_n)\}$ where the c_i is either 1 or −1 -- this constant denotes the class to which the point $\mathbf{x}_i$ belongs. Each $\mathbf{x}_i$ is a p – (statistics notation), or n – (computer science notation) dimensional vector of scaled [0,1] or [-1,1] values. The scaling is important to guard against variables (attributes) with larger variance that might otherwise dominate the classification. We can view this as *training data*, which denotes the correct classification which we would like the SVM to eventually distinguish, by means of the dividing hyperplane, which takes the form

Maximum-margin hyperplanes for a SVM trained with samples from two classes. Samples along the hyperplanes are called the support vectors.

Maximum-margin hyperplanes for a SVM trained with samples from two classes. Samples along the hyperplanes are called the support vectors.

$$\mathbf{w} \cdot \mathbf{x} - b = 0.$$

As we are interested in the maximum margin, we are interested in the support vectors and the parallel hyperplanes (to the optimal hyperplane) closest to these support vectors in either class. It can be shown that these parallel hyperplanes can be described by equations

$$\mathbf{w} \cdot \mathbf{x} - b = 1, \qquad (1)$$
$$\mathbf{w} \cdot \mathbf{x} - b = -1. \qquad (2)$$

We would like these hyperplanes to maximize the distance from the dividing hyperplane and to have no data points between them. By using geometry, we find the distance between the hyperplanes being $2/|w|$, so we want to minimize $|w|$. To exclude data points, we need to ensure that for all i either

$$\mathbf{w} \cdot \mathbf{x_i} - b \geq 1 \qquad \text{or}$$
$$\mathbf{w} \cdot \mathbf{x_i} - b \leq -1$$

This can be rewritten as:

$$c_i(\mathbf{w} \cdot \mathbf{x_i} - b) \geq 1 \quad 1 \leq i \leq n. \qquad (3)$$

The problem now is to minimize $|w|$ subject to the constraint (3). This is a quadratic programming (QP) optimization problem.

After the SVM has been trained, it can be used to classify unseen 'test' data. This is achieved using the following decision rule;

$$\hat{c} = \begin{cases} 1, & \text{if } \mathbf{w} \cdot \mathbf{x} + b \geq 0 \\ -1, & \text{if } \mathbf{w} \cdot \mathbf{x} + b \leq 0 \end{cases}$$

Writing the classification rule in its dual form reveals that classification is only a function of the Support vectors, i.e. the training data that lie on the margin.

Further theory

The use of the maximum-margin hyperplane is motivated by Vapnik Chervonenkis theory, which provides a probabilistic test error bound that is minimized when the margin is maximized. However the utility of this theoretical analysis is sometimes questioned given the large slack associated with these bounds: the bounds often predict more than 100% error rates.

The parameters of the maximum-margin hyperplane are derived by solving the optimization. There exist several specialized algorithms for quickly solving the QP problem that arises from SVMs. The most common method for solving the QP problem is Platt's SMO algorithm (http://research.microsoft.com/users/jplatt/smo.html) .

Non-linear classification

The original optimal hyperplane algorithm proposed by Vladimir Vapnik in 1963 was a linear classifier. However, in

1992, Bernhard Boser, Isabelle Guyon and Vapnik suggested a way to create non-linear classifiers by applying the kernel trick (originally proposed by Aizerman) to maximum-margin hyperplanes. The resulting algorithm is formally similar, except that every dot product is replaced by a non-linear kernel function. This allows the algorithm to fit the maximum-margin hyperplane in the transformed feature space. The transformation may be non-linear and the transformed space high dimensional; thus though the classifier is a hyperplane in the high-dimensional feature space it may be non-linear in the original input space.

If the kernel used is a Gaussian radial basis function, the corresponding feature space is a Hilbert space of infinite dimension. Maximum margin classifiers are well regularized, so the infinite dimension does not spoil the results. Some common kernels include,

- Polynomial (homogeneous): $k(\mathbf{x},\mathbf{x}') = (\mathbf{x} \cdot \mathbf{x}')^d$
- Polynomial (inhomogeneous): $k(\mathbf{x},\mathbf{x}') = (\mathbf{x} \cdot \mathbf{x}' + 1)^d$
- Radial Basis Function: $k(\mathbf{x},\mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2)$, for $\gamma > 0$
- Gaussian RBF: $k(\mathbf{x},\mathbf{x}') = \exp(-\frac{\|\mathbf{x} - \mathbf{x}'\|}{2\sigma^2})$
- Sigmoid: $k(\mathbf{x},\mathbf{x}') = \tanh(\kappa \mathbf{x} \cdot \mathbf{x}' + c)$, for some (not every) $\kappa > 0$ and $c < 0$

Soft margin

In 1995, Corinna Cortes and Vapnik suggested a modified maximum margin idea that allows for mislabeled examples. If there exists no hyperplane that can split the "yes" and "no" examples, the *Soft Margin* method will choose a hyperplane that splits the examples as cleanly as possible, while still maximizing the distance to the nearest cleanly split examples. This work popularized the expression *Support Vector Machine* or *SVM*. This method introduces slack variables and the equation (3) now transforms to

$$c_i(\mathbf{w} \cdot \mathbf{x_i} - b) \geq 1 - \xi_i \quad 1 \leq i \leq n \qquad (4)$$

This constraint in (4) along with the objective of minimizing $|w|$ can be solved using Lagrange multipliers or setting up a dual optimization problem to eliminate the slack variable.

Regression

A version of a SVM for regression was proposed in 1997 by Vapnik, Steven Golowich, and Alex Smola. This method is called support vector regression (SVR). The model produced by support vector classification (as described above) only depends on a subset of the training data, because the cost function for building the model does not care about training points that lie beyond the margin. Analogously, the model produced by SVR only depends on a subset of the training data, because the cost function for building the model ignores any training data that is close (within a threshold ε) to the model prediction.

References

- B. E. Boser, I. M. Guyon, and V. N. Vapnik. *A training algorithm for optimal margin classifiers*. In D. Haussler, editor, 5th Annual ACM Workshop on COLT, pages 144-152, Pittsburgh, PA, 1992. ACM Press.
- Corinna Cortes and V. Vapnik, "Support-Vector Networks, *Machine Learning, 20, 1995. [1] (http://citeseer.ist.psu.edu/cortes95supportvector.html)*
- Christopher J. C. Burges. "A Tutorial on Support Vector Machines for Pattern Recognition". Data Mining and Knowledge Discovery 2:121 - 167, 1998 *(Also available at CiteSeer: [2] (http://citeseer.ist.psu.edu/burges98tutorial.html))*
- Nello Cristianini and John Shawe-Taylor. *An Introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, 2000. ISBN 0-521-78019-5 *([3] (http://www.support-vector.net) SVM Book)*
- Huang T.-M., Kecman V., Kopriva I. (2006), Kernel Based Algorithms for Mining Huge Data Sets, Supervised,

Semi-supervised, and Unsupervised Learning, Springer-Verlag, Berlin, Heidelberg, 260 pp. 96 illus., Hardcover, ISBN 3-540-31681-7[4] (http://learning-from-data.com)

- Vojislav Kecman: "Learning and Soft Computing - Support Vector Machines, Neural Networks, Fuzzy Logic Systems", The MIT Press, Cambridge, MA, 2001.[5] (http://www.support-vector.ws)
- Tapio Pahikkala, Sampo Pyysalo, Jorma Boberg, Aleksandr Mylläri and Tapio Salakoski. Improving the Performance of Bayesian and Support Vector Classifiers in Word Sense Disambiguation using Positional Information. In Proceedings of the International and Interdisciplinary Conference on Adaptive Knowledge Representation and Reasoning (AKRR'05), Jun 2005.
- Bernhard Schölkopf and A. J. Smola: *Learning with Kernels*. MIT Press, Cambridge, MA, 2002. *(Partly available on line: [6] (http://www.learning-with-kernels.org) .)* ISBN 0-262-19475-9
- Bernhard Schölkopf, Christopher J.C. Burges, and Alexander J. Smola (editors). "Advances in Kernel Methods: Support Vector Learning". MIT Press, Cambridge, MA, 1999. ISBN 0-262-19416-3. [7] (http://www.kernel-machines.org/nips97/book.html)
- John Shawe-Taylor and Nello Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004. ISBN 0-521-81397-2 *([8] (http://www.kernel-methods.net) Kernel Methods Book)*
- P.J. Tan and D.L. Dowe (http://www.csse.monash.edu.au/~dld) (2004), MML Inference of Oblique Decision Trees (http://www.csse.monash.edu.au/~dld/David.Dowe.publications.html#TanDowe2004) , Lecture Notes in Artificial Intelligence (LNAI) 3339, Springer-Verlag, pp1082-1088 (http://www.csse.monash.edu.au/~dld/Publications/2004/Tan+DoweAI2004.pdf) . (This paper uses minimum message length (MML) and actually incorporates probabilistic support vector machines in the leaves of decision trees.)
- Vladimir Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, 1999. ISBN 0-387-98780-0

External links

- www.pascal-network.org (http://www.pascal-network.org) *EU Funded Network on Pattern Analysis, Statistical Modelling and Computational Learning*
- www.kernel-machines.org (http://www.kernel-machines.org) *(general information and collection of research papers)*
- www.kernel-methods.net (http://www.kernel-methods.net) *(News, Links, Code related to Kernel methods - Academic Site)*
- The Formulation of Support Vector Machine (http://svr-www.eng.cam.ac.uk/~kkc21/thesis_main/node8.html)
- www.support-vector.net (http://www.support-vector.net) *(News, Links, Code related to Support Vector Machines - Academic Site)*
- www.support-vector-machines.org (http://www.support-vector-machines.org) *(Literature, Review, Software, Links related to Support Vector Machines - Academic Site)*
- [9] (http://www.support-vector.ws) "(Free educational MATLAB based software for SVMs, NN and FL , Links, Publications downloads, Semisupervised learning software SemiL, Links)"

Software

- Lush (http://lush.sourceforge.net/) -- an amazing Lisp-like interpreted/compiled language with C/C++/Fortran interfaces that has packages to interface to a number of different SVM implementations. Interfaces to LASVM, LIBSVM, mySVM, SVQP, SVQP2 (SVQP3 in future) are available. Leverage these against Lush's other interfaces to machine learning, hidden markov models, numerical libraries (LAPACK, BLAS, GSL), and builtin vector/matrix/tensor engine.
- SVMlight (http://svmlight.joachims.org/) -- a popular implementation of the SVM algorithm by Thorsten Joachims; it can be used to solve classification, regression and ranking problems.
- LIBSVM -- A Library for Support Vector Machines, Chih-Chung Chang and Chih-Jen Lin (http://www.csie.ntu.edu.tw/~cjlin/libsvm/)
- YALE (http://yale.cs.uni-dortmund.de) -- a powerful machine learning toolbox containing wrappers for SVMLight, LibSVM, and MySVM in addition to many evaluation and preprocessing methods.
- LS-SVMLab (http://www.esat.kuleuven.ac.be/sista/lssvmlab) - Matlab/C SVM toolbox - well-documented, many features
- Gist (http://microarray.cu-genome.org/gist) -- implementation of the SVM algorithm with feature selection.

- Weka (http://www.cs.waikato.ac.nz/ml/weka/) -- a machine learning toolkit that includes an implementation of an SVM classifier; Weka can be used both interactively though a graphical interface or as a software library. (One of them is called "SMO". In the GUI Weka explorer, it is under the "classify" tab if you "Choose" an algorithm.)
- OSU SVM (http://sourceforge.net/projects/svm/) - Matlab implementation based on LIBSVM
- Torch (http://www.torch.ch/) - C++ machine learning library with SVM
- Spider (http://www.kyb.tuebingen.mpg.de/bs/people/spider/) - Machine learning library for Matlab
- e1071 (http://cran.r-project.org/src/contrib/Descriptions/e1071.html) - Machine learning library for R
- SimpleSVM (http://asi.insa-rouen.fr/~gloosli/simpleSVM.html) - SimpleSVM toolbox for Matlab
- SVM and Kernel Methods Matlab Toolbox (http://asi.insa-rouen.fr/~arakotom/toolbox/index.html)
- PCP (http://pcp.sourceforge.net/) -- C program for supervised pattern classification. Includes LIBSVM wrapper.

Interactive SVM applications

- ECLAT (http://mips.gsf.de/proj/est/) classification of Expressed Sequence Tag (EST) from mixed EST pools using codon usage
- EST3 (http://mips.gsf.de/proj/este/) classification of Expressed Sequence Tag (EST) from mixed EST pools using nucleotide triples

Retrieved from "http://en.wikipedia.org/wiki/Support_vector_machine"

Categories: Machine learning | Classification algorithms | Ensemble learning

---